

Naturalment

Article

Projecte Genoma Humà

Història del Projecte del Genoma Humà

Dècada dels anys 80, un espai temporal d'inflexió en molts aspectes de la humanitat, ja sigui a nivell social com a nivell tecnològic / científic. És a partir d'aquest moment en què podem començar a parlar del PGH, Projecte Genoma Humà.

A mitjans d'aquesta dècada, els genetistes van començar a plantejar la idea de poder seqüenciar els aproximadament 3×10^9 parells de bases del genoma humà. Cal dir que dins d'aquesta època ja hi va haver una gran quantitat d'experiments en aquest àmbit, però centrats en la seqüenciació de gens individuals d'una gran diversitat d'organismes. En canvi, amb aquest projecte es volia seqüenciar el màxim possible el genoma humà.

Va ser als Estats Units, l'any 1986, quan es va concretar la creació del Projecte Genoma Humà, quan el Departament d'Energia (Department of Energy), en un congrés a Santa Fe, va plantejar la idea de destinar diners a la investigació i seqüenciació del genoma humà, amb l'objectiu de poder lluitar contra les radiacions que podia haver-hi en sobre del material genètic hereditari i avançar en diversos àmbits de la ciència.

Durant l'any següent, diverses institucions es van unir a la idea proposada pel Departament d'Energia. Fins que l'any 88, un any després, es va crear l'organització HUGO (Human Genome Organization). Aquesta organització era a nivell mundial, la qual tenia l'objectiu de col·laborar entre una gran quantitat de països en la investigació del genoma humà, perquè així no hi hagués redundàncies i poder difondre més fàcilment els avenços dins d'aquest experiment. Aquest projecte va ser aprovat pels Estats Units l'any 90, els quals van aportar 3.000 milions de dòlars, amb la condició que en 15 anys estigués completament seqüenciat el genoma humà.

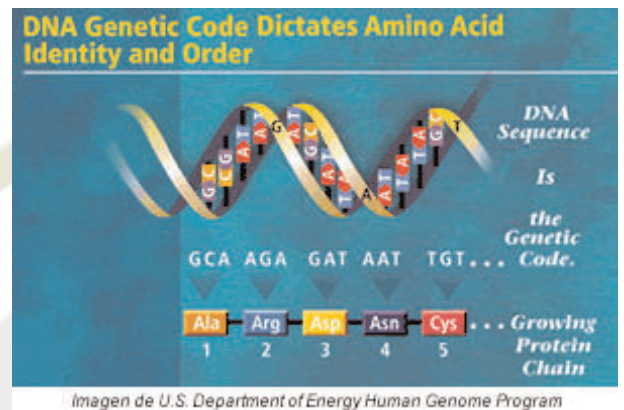
Cal dir, que el projecte no només tenia com a objectiu la seqüenciació del genoma, sinó també la creació d'infraestructures tecnològiques, com

per exemple trobar un mètode per facilitar el processament d'tantes dades i també, la identificació dels gens.

A més, hi havia objectius com: la creació de mapes genètics, la determinació de seqüències completes de diverses espècies, com ara l'*Escherichia coli* o la *Drosophila melanogaster*, i, sobretot, la creació de tecnologies dins de l'ambient biològic i informàtic, és a dir, una millora en el camp de la bioinformàtica.

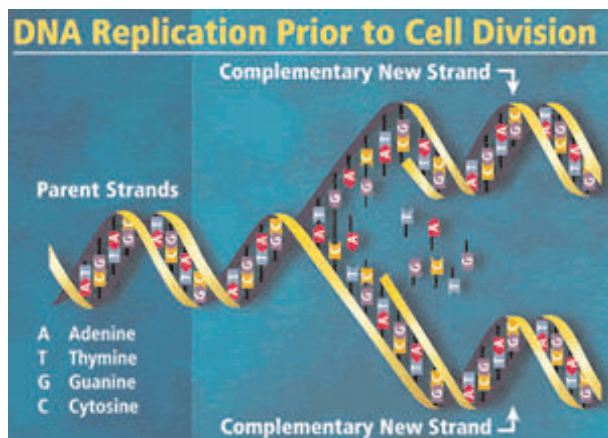
El Projecte Genoma Humà va començar oficialment a l'octubre de 1990. El primer objectiu era aconseguir desenvolupar nous mètodes automatitzats per clonar i seqüenciar l'ADN, i així generar detallats mapes físics i genètics del genoma humà. No va ser fins tres anys després quan es van completar els mapes físics a gran escala dels 23 cromosomes humans.

Mentre que el Consorci Internacional de Seqüenciació del Genoma Humà va utilitzar un



esquema basat en els mapes físics per seqüenciar el genoma, un procés que comportava un gran esforç i es necessitava molt de temps, Craig Venter, un reconegut biòleg i empresari nord-americà, va decidir el 1998 dirigir una empresa privada anomenada Celera Genomics, la qual es creava amb l'objectiu de seqüenciar el genoma humà abans que el Consorci.

Per dur a terme la seva investigació, Craig Venter va utilitzar un esquema de seqüenciació per fragments escollits a l'atzar, aquest procés es va denominar Shotgun. Amb això, volia aconseguir arribar a seqüenciar el genoma humà abans que el Consorci.



Imatge de U.S. Department of Energy Human Genome Program

Durant molts anys, tant el Consorci com l'empresa Celera van anar avançant de forma simultània en la investigació. Però no va ser fins a l'estiu de l'any 2000, quan, tant l'organització pública com l'empresa privada, van anunciar que ja estava complet el primer esborrany, el qual incloïa la major part de la seqüència del genoma humà.

No obstant això, no va ser fins a la primavera de 2003 quan la seqüència del genoma humà va ser declarada completa, tot i que encara queden alguns escletxes per completar.

Arran dels projectes Genoma Humà (PGH), creada per HUGO (Human Genome Organization), i Celera Genomics, creada per Craig Venter, es van presentar dos primers esborranys el 2001, la comparació dels quals revelava que el mètode utilitzat pel consorci internacional de clonació era més precís que l'empleat per Celera Genomics. Tanmateix, el mètode del shotgun era molt més ràpid pel que permetia obtenir uns primers resultats molt abans. Les dades més significatives extretes d'aquests primers esborranys van ser:

- El nombre estimat de gens es comprenia entre 30.000 i 35.000, davant dels 150.000 gens que s'estimava en un principi que hauria de constar el genoma humà.
- Al voltant del 60% dels gens presentaven homologies amb altres genomes seqüenciats prèviament.
- Únicament 94 de les 1.278 famílies de gens previstes eren específiques d'animals vertebrats.

el qual indica que comptem amb més similituds amb organismes invertebrats.

- Els exons (regions de l'ADN que donen lloc a proteïnes) ocupaven només el 1,1% del genoma, mentre que els introns (regions de l'ADN que separen dues regions d'exons i que no donen lloc a proteïnes) el 24%. El 75% restant era ADN intergènic i el 44% del mateix derivaria d'elements transposables.

Tres anys després, HUGO publica la seqüència pràcticament definitiva del 99% de les regions eucariòtiques del genoma humà. En aquesta, els autors van estimar el nombre total de gens que codifiquen proteïnes entre els 20.000 i 25.000. Això suposa un nombre fins i tot menor que l'obtingut en el primer esborrany, i que deixo perplexos els investigadors.

A causa de la complexitat que suposadament se li atribueix a l'ésser humà, es pressuposava que el nombre de gens que presentaria havia d'estar directament relacionat amb la mateixa i per tant presentar un gran nombre (sobre els 150.000). Tanmateix, és inevitable la perplexitat al 80 trobar-nos que només consta d'uns 20.000 o 25.000 gens, que en comparar-los amb els genomes d'altres espècies seqüenciades, com *Mus Musculus* (ratolí) amb 22.705 gens, *Arabidopsis thaliana* (Planta crucífera) amb 27.416, trobem que contenen més gens.

Organisme genètic model

Les eines creades pel PGH també seguiran temptant els esforços per caracteritzar els genomes complets de diversos altres organismes que s'utilitzen àmpliament en la investigació biològica, com ara ratolins, mosques i cucs plans. Aquests esforços es recolzen mútuament, perquè la majoria dels organismes tenen molts gens similars, o "homòlegs", amb funcions similars. Per tant, la identificació de la seqüència o la funció d'un gen en un organisme model té el potencial per explicar un gen homòleg en els éssers humans, o en un dels altres organismes model.

La mosca de la fruita: *Drosophila melanogaster*

Aquesta mosca va ser un dels primers organismes usats per l'anàlisi genètica i actualment és un dels organismes eucariotes més utilitzats i millor coneguts des del punt de vista genètic. Atès que tots els organismes utilitzen un sistema genètic comú, el coneixement d'un procés com la reproducció o la transcripció en les mosques de la fruita ens ajuda a comprendre aquests mateixos processos en els éssers humans.

Drosophila és un gènere de més de 1.000 espècies descrites de mosques petites que sovint s'alimenten i es reproduïen en la fruita, encara que poques vegades provoca mal i no són considerades plagues econòmiques. La mosca de la fruita millor coneguda i més estudiada és *D. melanogaster*.

D. melanogaster començar a aparèixer en els laboratoris biològics al voltant de 1900. Thomas Hunt Morgan va començar a utilitzar les mosques de la fruita en estudis experimentals d'herència a la Columbia University.

La mosca de la fruita té algunes característiques que la converteixen en un subjecte ideal per a les investigacions genètiques:

- Posseeix un temps de generació relativament curt; completen una generació sencera en aproximadament 10 dies a temperatura ambient.
- Encara que curt, posseeix un cicle vital complex, que travessa diversos estadis evolutius diferents, que inclouen ou, larva, pupa i adult.
- Produeixen gran quantitat de descendents; posen de 400 a 500 ous en un període de 10 dies.
- Són fàcils de criar al laboratori. Es crien en petits flascons de vidre o ampolles amb un aliment similar a una pasta que consisteix en plàtans o un cereal i melassa.
- Els mascles i les femelles es distingeixen fàcilment i les femelles verges s'aïllen amb facilitat, el que facilita els encreuaments genètics.
- Les mosques són petites i requereixen poc espai, alhora són prou grans com perquè es puguin observar amb facilitat moltes mutacions amb una lent portàtil o un microscopi de dissecció.
- *D. melanogaster* té un genoma relativament petit que consisteix en 175 milions de parells de bases de DNA, el que és només al voltant del 5% del genoma humà. Té quatre parells de cromosomes: tres parells d'autosomes i un parell de cromosomes sexuals.

L'any 2000 es va seqüenciar el genoma complet de *D. melanogaster*. *Drosophila* continua sent un dels organismes genètics model més versàtils i poderosos.

El bacteri *Escherichia coli*

És el microorganisme procariota estudiat amb més profunditat i un dels millors caracteritzats des del punt de vista genètic de totes les espècies. Aquest bacteri resideix de manera natural al

tracte intestinal dels éssers humans i d'altres animals de sang calenta. *E. coli* va ser descrita per primera vegada per Theodore Escherich en 1885, però durant molts anys es va suposar que tots els bacteris només es reproduïen en forma asexual i que els encreuaments genètics eren impossibles. En 1946 Joshua Lederberg i Edward Tatum van demostrar que *E. coli* pateix un tipus de reproducció sexual.

Avantatges d'*E. coli* com un microorganisme model:

- La seva doble avantatge és la ràpida multiplicació i la petita grandària. En condicions òptimes, aquest microorganisme pot reproduir cada 20 minuts i, en amb prou feines 7 hores, una cèl·lula bacteriana sola pot donar origen a més de 2 milions de descendents.
- És fàcil de conrear al laboratori en medis líquids o en medis sòlids en plaques de Petri.

El genoma d'*E. coli*

El genoma d'*E. coli* està en un sol cromosoma i, en comparació amb el dels éssers humans, els ratolins, i altres organismes multicel·lulars, és relativament petit; consta de poc més de quatre milions i mig de parells de bases. S'estima que el genoma d'*E. coli* conté 4.300 gens, dels quals més de la meitat no té cap funció coneguda. El genoma haploide de *E. coli* facilita l'aïllament de mutacions perquè no hi ha gens dominants en el mateix locus que suprimeixin i emmascarin les mutacions recessives.

Escherichia coli s'utilitza en diversos sistemes d'experimentació en què s'estudien en detall els processos genètics fonamentals. Els sistemes in vitro basats en components provinents de cèl·lules *E. coli* permeten que s'estudiïn i analitzin la transcripció, la replicació, l'expressió gènica i moltes altres funcions genètiques importants sota condicions de laboratori controlades.

E. coli també s'utilitza molt en enginyeria genètica.

La planta *Arabidopsis thaliana*

La major part dels treballs inicials en genètica es van realitzar en plantes, però, cap a mitjan segle XX aquestes es van deixar de banda. *A. thaliana* va ser identificada al segle XVI, però l'estudi d'aquesta espècie no es va generalitzar fins que van aparèixer els primers mapes genètics detallats a principis de la dècada de 1980.

Avantatges d'*Arabidopsis* com a organisme genètic model:

- Té una mida petita: l'alçada màxima que assoleix és d'entre 10 i 20 cm.

- Reproducció prolífica
- Petit genoma

Arabidopsis thaliana completa el seu desenvolupament, des de la germinació de la llavor fins a la producció d'aquesta, al voltant de 6 setmanes. La seva petita grandària i la seva capacitat per créixer amb escassa llum la fan ideal per al cultiu al laboratori. Cada planta és capaç de produir entre 10.000 i 40.000 llavors, amb una alta taxa de germinació.

El petit genoma d'*A. thaliana* té només 125 milions de parells de bases de DNA en 5 parells de cromosomes, en comparació amb els 2.500 milions de parells de bases del DNA del genoma del blat de moro i els 16.000 milions de parells de bases en el genoma del blat. El genoma d'*A. thaliana* es va seqüenciar per complet l'any 2000.

El cuc nematode *Caenorhabditis elegans*

Si bé se'ls veu en rares ocasions, els nematodes són dels organismes més abundants al nostre planeta a causa que habiten els sòls de tot el món. S'ha usat extensament en estudis genètics arran del seu cos pla simple, la seva facilitat per conrear-lo i la seva elevada capacitat reproductiva. Va ser introduït inicialment en els estudis genètics per Sydney Brenner. Avantatges de *C. elegans* com a organisme genètic model: Com els anteriors, el cuc nematode és petit, fàcil de conrear i produeix una progènie 180 nombrosa.

- El *C. elegans* adult mesura al voltant d'1 mm de llarg. La majoria dels investigadors el conreen en plaques de Petri amb agar, entapissades amb una gaspa de bacteris, que els nematodes devoren.
- Comparats amb la majoria dels animals multicel·lulars, tenen un temps de generació curt, al voltant de 3 dies a temperatura ambient.
- D'altra banda, són prolífics, ja que una sola femella és capaç de produir entre 250 i 1.000 ous fecundats en 3 a 4 dies.
- És transparent. Aquesta característica permet l'observació del desenvolupament intern en totes les etapes.

La major part dels adults madurs són hermafrodites, amb la capacitat de produir òvuls i espermatozoides, i d'autofecundar. Uns pocs són mascles.

Els òvuls es fecunden internament. Després es produeix la fresa i el desenvolupament es completa de manera externa. Al voltant de 14 hores després de la fecundació, eclosiona l'ou del qual sorgeix una

larva, que travessa 4 estats larvaris (L1, L2, L3, L4) separats per mudes. En condicions normals de laboratori, els cucs viuen entre 2 i 3 setmanes.

El *Caenorhabditis elegans* té 5 parells d'autosomes més 2 cromosomes X en les femelles (hermafrodites) o 1 cromosoma X en els mascles. La seva DNA té 103 milions de parells de bases i el nombre de gens és de 20.500. El genoma de *C. elegans* va ser seqüenciat l'any 1998.

El llevat *Saccharomyces cerevisiae*

A partir de l'aïllament i l'estudi de les mutacions petites, el llevat comú de forn ha estat utilitzat àmpliament per a l'estudi dels sistemes genètics mitocondrials. Louis Pasteur va identificar a *S. cerevisiae* com el microorganisme responsable de la fermentació en 1857 i el seu ús en l'anàlisi genètica va començar aproximadament a 1935.

Avantatges del llevat com a organisme genètic model:

- És un organisme eucariota, amb sistemes genètic i cel·lular similars als d'altres eucariotes més complexos, com els éssers humans.
- És unicel·lular.
- Les cèl·lules de llevat necessiten poc espai i gran quantitat de cèl·lules poden créixer fàcilment i amb baix cost en el laboratori.
- El llevat existeix en formes diploides i haploides. Quan són haploides l'al·lel serà expressat en el fenotip. Per tant, es podran identificar fàcilment els al·lels recessius en les cèl·lules haploides i després es podran examinar les interaccions entre al·lels en les cèl·lules diploides.
- Després de la meiosi, tots els productes d'una divisió meiótica són presents en una única estructura anomenada "fàstic" i es mantenen separats dels productes d'altres divisions meiótiques. Les quatre cèl·lules produïdes per una única divisió meiótica s'anomenen tètada. El fet de comptar amb tètades separades en el llevat permet observar directament els efectes de les divisions meiótiques individuals sobre els tipus de gàmetes produïts i identificar més fàcilment els esdeveniments d'entrecreuament. L'anàlisi genètica d'una tètada s'anomena anàlisi de la tètada.
- L'estudi genètic de les cèl·lules del llevat sovint contribueix al nostre coneixement d'altres organismes eucariotes més complexos, inclosos els éssers humans.

Saccharomyces cerevisiae té 16 parells de cromosomes eucariotes típics. El genoma de *S. cerevisiae* conté 12 milions de parells de bases. El 1996, *S. cerevisiae* va ser el primer organisme eucariota el genoma es va seqüenciar totalment.

El ratolí *Mus musculus*

El ratolí comú domèstic figura entre les espècies més antigues i valuoses per als estudis genètics. És un organisme genètic excel·lent: petit, prolífer i fàcil de mantenir, amb un temps de generació curt.

Avantatges del ratolí com un organisme genètic model:

- Estreta relació evolutiva amb els éssers humans. Per ser un mamífer, el ratolí és des del punt de vista genètic, conductual i fisiològic, més similar als éssers humans que altres organismes usats en els estudis de genètica, el que el converteix en el model d'elecció per a molts estudis de genètica humana i mèdica.
- Té un temps de generació curt, en comparació amb el de la majoria dels altres mamífers.
- *Mus musculus* es conserva i cria en gàbies que requereixen poc espai al laboratori.
- Els ratolins tenen ventrades grans (8 a 10 cries) i són dòcils i fàcils de manejar.

La producció de gàmetes i la reproducció en el ratolí són molt similars a les dels éssers humans. El període de gestació en ratolins dura generalment 21 dies. Aquests arriben a la pubertat en unes 5 a 6 setmanes i poden viure al voltant de 2 anys.

El genoma del ratolí conté al voltant de 2.600 milions de parells de bases de DNA, el que és similar en grandària al genoma humà. Aquest genoma està distribuït en 19 parells de cromosomes autosòmics i un parell de cromosomes sexuals.

El genoma humà

El PGH ha revelat que hi ha probablement prop de 20.500 gens humans. La seqüència humana completa ara pot identificar les seves ubicacions. Aquest últim producte del PGH ha donat al món una font d'informació detallada sobre l'estructura, organització i funcionament de tot el conjunt de gens humans. Aquesta informació pot ser pensada com el conjunt bàsic per al desenvolupament i la funció d'un ésser humà.

El genoma humà s'ha estudiat i analitzat extensament causa de la seva importància per a la salut i l'evolució. Té una longitud de 3.200 milions

de parells de bases, però només el 25% del DNA es transcriu a RNA, i menys del 2% codifica proteïnes. La grandària mitjana dels gens en el genoma humà és d'aproximadament 27.000 parells de bases, amb 9 exons. Un mateix gen sol codificar diverses proteïnes gràcies al tall i entroncament alternatiu; cada gen codifica, de mitjana, dues o tres mRNA diferents, el que implica que el genoma humà, amb aproximadament 24.000 gens, podria codificar 72.000 proteïnes o més.

Podem arribar a la conclusió, per tant, que hi ha un alt percentatge de DNA que no codifica proteïnes ni es transcriu a RNA. A aquest DNA se li va cridar "DNA escombraries" (junk DNA) ja que es va pensar que no tenia cap funció més que la d'acumular mutacions de manera que així el veritable ADN codificant, en aquell moment l'important, el que porta la informació per formar les proteïnes, estigués protegit

Bibliografia

- Bergel, S.D. El genoma humano y los límites del patentamiento. *Interciencia.org*. Consultado [Febrero 2013].
- Carroll, M.C., Ciaffa, J. El proyecto genoma humano: una revisión científica y ética. *ActionBioscience.org*. Consultado [Febrero 2013]. Disponible en (<http://www.actionbioscience.org>)
- Mural, R.J., et al. (2002). A Comparison of Whole-Genome Shotgun-Derived Mouse Chromosome 16 and the Human Genome. *Science*. Vol. 296 no. 5573. 1661-1671. DOI: 10.1126/science.1069193
- The 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491. 56-65. DOI:10.1038/nature11632. Disponible en (<http://www.nature.com>)